

SPECIFICATION AMENDMENTS

[0001] This invention is related generally to speaker independent [[voice]] speech recognition (SIVR), and more specifically to speech-enabled applications using dynamic context switching and multi-pass parsing during speech recognition.

[0004] The problems with the existing speech recognition engines, mentioned above, have prevented a speech-enabled user interface from becoming a practical alternative to data entry and operation of information displays using short command phrases. True speaker independent [[voice]] speech recognition (SIVR) is needed to make a speech-enabled user interface practical for the user.

[0005] Pre-existing [[SIVR]] systems like the one marketed by Fluent Technologies, Inc. can only be used with limited vocabularies, typically 200 words or less, in order to keep recognition error rates acceptably low. As the size of a vocabulary increases, the recognition rate of a speech engine decreases, while the time it takes to perform the recognition increases. Some applications for speech-enabled user interfaces require a vocabulary several orders of magnitude larger than the capability of Fluent's engine. Applications can have vocabularies of 2,000 to 20,000 words that must be handled by the [[SIVR]] system. Fluent's speech recognition engine is typically applied to recognize short command phrases, with a command word and one or more command parameters. The existing approach to parsing these structured sentences, is to first express the recognition context as a grammar that encompasses all possible permutations and combinations of the command words and their legal parameters. However, with long command sentences and/or with "non-small" vocabularies for the modifying parameters ("data rich" applications), the number of permutations and combinations increases beyond the speech engine's capability of generating unambiguous results. Existing [[SIVR]] systems, like the Fluent system discussed herein are inadequate to meet the needs of a speech-enabled user interface coupled to a "data rich" application.

[0006] What is needed is a [[SIVR]] system that can translate a long command phrase and/or a “non-small” vocabulary for the modifying parameters, with high accuracy in real-time.

[0011] **Figure 4** provides a general system architecture that achieves achieving speaker independent [[voice]] speech recognition.

[0017] A system architecture is disclosed for designing a speech-enabled user interface of general applicability to a subset of a language vocabulary. In one or more embodiments, the system architecture, multi-pass parsing, and dynamic context switching are used to achieve speaker independent [[voice]] speech recognition (SIVR) of a speech-enabled user interface. The techniques described herein are generally applicable to a broad spectrum of subject matter within a language vocabulary. The detailed description will flow between the general and the specific. Reference will be made to a medical subject matter during the course of the detailed description, no limitation is implied thereby. Reference is made to the medical subject matter to contrast the general concepts contained within the invention with a specific application to enhance communication of the scope of the invention.

[0026] The preceding general description is contained within the block diagram of **Figure 4** at 400. **Figure 4** provides a general system architecture that achieves speaker independent [[voice]] speech recognition by combining the methodology according to the teaching of the invention. A subset of a language vocabulary is defined for translating speech into text at block 402. The subset is separated into a plurality of contexts at block 404. A speech signal is divided between a plurality of contexts at block 406. A set of constraint filters is applied to a plurality of contexts at block 408. Speech recognition is performed on the speech signal using multi-pass parsing at block 410. The speech recognition is biased using constraint filters at block 412. Contexts are dynamically switched during speech recognition at block 414. In various embodiments, the general principles contained in **Figure 4** are applicable to wide variety of subject matter as previously discussed. These general principles may be used to design applications using a speech-enabled user interface. In one embodiment, **Figure 5** illustrates a flow chart depicting a

process for building a speech-enabled user interface for a medical application. With reference to **Figure 5**, a user interface for a speech enabled medical application is defined at block 502. Block 502 includes designing screens for the medical application and speech-enabled input fields. A vocabulary associated with each input field is defined at block 504. The associated constraint filters are defined at block 506 for the medical setting. Blocks 502, 504, and 506 come together at block 508 to provide an application that constrains the language vocabulary during run-time of the application, utilizing the speech engine to convert speech to text independent of the speaker's [[voice]] speech. In one embodiment, the present invention is producing 95% accurate identification of speech with vocabularies of over 2,000 words. This is a factor of 10 improvement in vocabulary size, for the same accuracy rating, over existing speech identification techniques that do not utilize the teachings of the present invention.

[0031] Many other business applications are contemplated. A non-exclusive list includes business entities such as an automotive company, a financial services company, a bank, an investment company, an accounting firm, a law firm, a grocery company, and a restaurant services company. In one embodiment, a business entity will receive the signal resulting from the speech recognition process according to the teachings of the present invention. In one embodiment, the user of the speech-enabled user interface will be able to interact with the business entity using the handheld device with [[voice]] speech as the primary input method. In another embodiment, a vehicle, such as a car, truck, boat or air plane, may be equipped with the present invention allowing the user to make reservations at a hotel or restaurant or order a take-out meal instead. In another embodiment, the present invention may be an interface within a computer (mobile or stationary).

[0033] Thus, a novel speaker independent [[voice]] speech recognition system (**SIVR**) is described. Although the invention is described herein with reference to specific preferred embodiments, many modifications therein will readily occur to those of ordinary skill in the art. Accordingly, all such variations and modifications are included within the intended scope of the invention as defined by the following claims.